The case for caution in predicting scientists' future impact

Orion Penner,[1] Raj K. Pan,[2] Alexander M. Petersen,[3] and Santo Fortunato[2]

[1]*Laboratory of Innovation Management and Economics, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy*
[2]*Department of Biomedical Engineering and Computational Science, Aalto University School of Science, P.O. Box 12200, FI-00076, Finland*
[3]*Laboratory for the Analysis of Complex Economic Systems, IMT Lucca Institute for Advanced Studies, 55100 Lucca, Italy*

We stress-test the career predictability model proposed by Acuna et al. [Nature 489, 201-2 2012] by applying their model to a longitudinal career data set of 100 Assistant professors in physics, two from each of the top 50 physics departments in the US. The Acuna model claims to predict $h(t + \Delta t)$, a scientist's h-index Δt into the future, using a linear combination of 5 cumulative career measures taken at career age t. Here we investigate how the "predictability" depends on the aggregation of career data across multiple age cohorts. We confirm that the Acuna model does a respectable job of predicting $h(t + \Delta t)$ up to roughly 6 years into the future when aggregating all age cohorts together. However, when calculated using subsets of specific age cohorts (e.g. using data for only $t = 3$), we find that the model's predictive power significantly decreases, especially when applied to early career years. For young careers, the model does a much worse job of predicting future impact, and hence, exposes a serious limitation. The limitation is particularly concerning as early career decisions make up a significant portion, if not the majority, of cases where quantitative approaches are likely to be applied.

arXiv:1304.0627v1 [physics.soc-ph] 2 Apr 2013

Any scientist pursuing a research career these days is acutely aware of the increasingly central role metrics play in measuring scientific impact. From papers to people, the quality of almost everything is being measured by citations. One area in which metrics are starting to cause a shift is in the scientific career evaluation process. From a purely economic point of view, a tenure track hire is a million dollar bet on a young scientist's future success, so it is easy to see why metrics and models capable of predicting future success are very attractive to decision makers, but it also highlights that this "genie" is unlikely to be put back in its bottle.

If metrics are going to be integrated into the career advancement process they must be better tested and many specific questions need to be investigated. For example, what aspects of a career are *actually* predictable? What ingredients are required for a model to be robust? How often is a given model's prediction wrong, and what impact does that have on the careers of scientists, especially young ones that are already burdened by risk [1]? Without a proper understanding of the above questions, any uncritical use of quantitative indicators can do real harm to scientists and to the endeavor of science as a whole.

The introduction of the *h*-index [2] in late 2005 was a significant milestone in the use of metrics in career evaluation. Figure 1(A) shows that the popularity of the h-index has consistently increased since its introduction. Now it stands as the most popular quantitative measure for a researcher's productivity and impact. In fact, it is already being used to evaluate scientists, e.g., a modified version of the *h*-index has been integrated into the Italian national tenure competition [3].

Of course future impact, rather than past accomplishments,

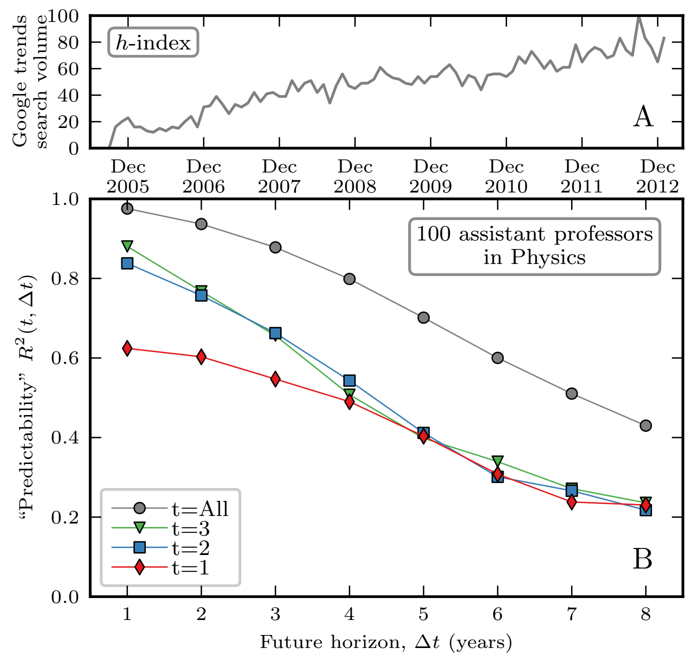

FIG. 1: (A) Google search volume, normalized to % peak value, is a proxy for the interest in the "h-index". (B) The "predictive power" of the Acuna model [5] decreases significantly when early career age cohorts (years since first publication $t = 1, 2, 3$) are analyzed separately.

is really what's at the heart of most career appraisal decisions in science, e.g., tenure, grants, fellowships, prizes, etc. So how predictable is an individual's future *h*-index? Previously, it has been indicated that the *h*-index is better than other indicators in predicting future scientific achievements [4]. A more recent publication by Acuna *et. al.* [5] presents a model that predicts an individual's future *h*-index using a linear combination of five other metrics. In its technical details this work is

[1] Published in *Physics Today* **66**, 8–9 (2013).
Send correspondence to:
petersen.xander@gmail.com or santo.fortunato@aalto.fi

notable because it is one of the first to integrate several metrics into a prediction. However it is probably more noteworthy in its non-technical aspects given the high profile forum in which it was published and the authors' suggestion that it can be used in decision making, going as far as to provide an online future h-index calculator.

In the Acuna *et. al.* model, an individual's future h-index, $h(t + \Delta t)$, is modeled as a linear combination of his/her (i) current h-index $h(t)$, (ii) square root of number of publications $\sqrt{N}$, (iii) number of years since first publication t, (iv) number of publications in high impact journals q, and (v) number of distinct journals j. By incorporating several key aspects of academic publishing their multiple regression model appears quite promising. However further investigation highlights the care that needs to be taken in developing models of future impact.

To illustrate the difficulties of predicting future success we applied the Acuna model to a longitudinal career data set of 100 Assistant professors in physics, two from each of the top 50 physics departments in the US (see [1] for further description of this dataset). Figure 1(B) shows the coefficient of determination R^2 for the regression model Δt years into the future using data available at "career age" t. The Acuna model aggregates all years in the data sample together ($t =$ All, black curve), and in doing so it yields a respectable prediction of $h(t + \Delta t)$ even up to $\Delta t = 6$ years. However, we find that the model's predictive power depends strongly on the mixing of the age cohorts.

To demonstrate the model's dependence on mixing of career ages, we also show the Acuna model $R^2(t, \Delta t)$ calculated without aggregating data across all t (colored curves). From this one can clearly see that the $R^2(t, \Delta t)$ values calculated for a fixed t are significantly less than the $R^2(\Delta t)$ values calculated by aggregating across all career ages. This means the model is generally poor at predicting the future success of early career scientists. We also note that artificially large R^2 values can follow from predictability models which use cumulative variables (e.g. $h(t)$ which is non-decreasing), as opposed to incremental variables, such as $\Delta h(t, \Delta t) \equiv h(t + \Delta t) - h(t)$ [6, 7]. These limitations are particularly concerning as early career decisions make up a significant portion of cases where quantitative approaches are likely to be applied. We further confirmed our observation of much lower R^2 values in the early career (t up to 5 years) using additional career data for 200 highly cited physicists.

Recent work by A. Mazloumian hints at one of the underlying difficulties of predicting a scientist's future success [8]. By differentiating between citations accrued by papers *already published* at the time of prediction and citations accrued by future papers *published after* the prediction time, it is shown that regression approaches do a reasonable job predicting future citations to past papers, but are unable to reliably predict *future* citations to *future* papers. In the context of predicting the future impact of a scientist, this means that there is not necessarily a correlation between the impact of papers published in the past and the impact of papers published in the future.

Going forward, these approaches and their successors will be increasingly exploited in real decision making processes. However, at the moment little is known about the strengths and weaknesses of the state-of-the-art predictive indicators. It is open to debate with whom the *responsibility* to vet current and new quantitative measures lies. But what is clear is that scientists themselves, particularly young ones, stand to lose the most should quantitative measures be stretched too far in the realm of career decisions. As a community it behooves us to engage with the institutions that seek to exploit quantitative measures of scientific impact in their decision making process, while maintaining a skepticism backed by quantitative and rigorous analysis of the specific measures they seek to employ.

[1] Petersen, A. M., Riccaboni, M., Stanley, H. E., Pammolli, F. Persistence and Uncertainty in the Academic Career. *Proc. Nat. Acad. Sci. USA* **109**, 5213 – 5218 (2012).

[2] Hirsch, J. E. An index to quantify an individual's scientific research output. *Proc. Natl. Acad. Sci. USA* **102**, 16569–16572 (2005).

[3] National Agency for the Evaluation of Universities and Research Institutes (Italy), www.anvur.org/sites/anvur-miur/files/normalizzazione_indicatori_0.pdf

[4] Hirsch, J. E. Does the h index have predictive power? *Proc. Natl. Acad. Sci.* **104**, 19193–19198 (2008).

[5] Acuna, D. E., Allesina, S., Kording, K. P. Future impact: Predicting scientific success. *Nature* **489**, 201–202 (2012).

[6] Granger, C. W. J., Newbold, P. Spurious regressions in econometrics. *J. Econometrics* **2**, 111–120 (1974).

[7] O. Penner, R. K. Pan, A. M. Petersen, S. Fortunato. Submitted (2013).

[8] Mazloumian, A. Predicting Scholars' Scientific Impact. *PLoS ONE* **7**, (2012).